Project 2

Daniel Fredin, Junhan Li, & Eric Chen

Problem 1

Problem 3

a)

ii) 10-fold Cross-Validation

Model Accuracy

```r
# Predict outcome using model from training data based on testing data
predictions <- predict(cvmodel, newdata=test)

# Create confusion matrix to assess model fit/performance on test data
con_matx <- confusionMatrix(data=predictions, test$COMMIT)
con_matx
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction no yes
##        no  13   4
##        yes  2  11
##
##                Accuracy : 0.8
##                  95% CI : (0.6143, 0.9229)
##     No Information Rate : 0.5
##     P-Value [Acc > NIR] : 0.0007155
##
##                   Kappa : 0.6
##
##  Mcnemar's Test P-Value : 0.6830914
##
##             Sensitivity : 0.8667
##             Specificity : 0.7333
##          Pos Pred Value : 0.7647
##          Neg Pred Value : 0.8462
##              Prevalence : 0.5000
##          Detection Rate : 0.4333
##    Detection Prevalence : 0.5667
##       Balanced Accuracy : 0.8000
```

```
##
##        'Positive' Class : no
##
```

b)

c)

```r
compute_weibull_stats <- function(shape, scale) {
  # Compute the mean
  mean_value <- scale * gamma(1 + (1/shape))


  # # Compute the median
  # median_value <- scale * qweibull(0.5, shape, scale)


  # Compute the median
  median_value <- scale * (log(2)^(1/shape))


  # Compute the mode
  if (shape > 1)
    mode_value <- scale * ((shape - 1) / shape)^(1/shape)
  else
    mode_value <- 0

  # Compute the variance
  variance_value <- scale^2 * (gamma(1 + (2/shape)) - (gamma(1 + (1/shape)))^2)

  # Return the computed statistics as a named list
  return(list(mean = mean_value,
              median = median_value,
              mode = mode_value,
              variance = variance_value))
}

# lambda = scale
# K = shape

# Example usage
scale_param <- 6
shape_param <- 1

stats <- compute_weibull_stats(shape_param, scale_param)

# Accessing the computed statistics
mean_value <- stats$mean
median_value <- stats$median
mode_value <- stats$mode
variance_value <- stats$variance
```

```r
# Printing the computed statistics
cat("Mean:", mean_value, "\n")
```

```
## Mean: 6
```

```r
cat("Median:", median_value, "\n")
```

```
## Median: 4.158883
```

```r
cat("Mode:", mode_value, "\n")
```

```
## Mode: 0
```

```r
cat("Variance:", variance_value, "\n")
```

```
## Variance: 36
```

Maximum likelihood estimator (MLE) of the parameters from the Weibull distribution

Probability Density Function ($f(x; \lambda, k)$) for the Weibull distribution:

$$f(x; \lambda, k) = \begin{cases} \frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{k-1} e^{-\left(\frac{x}{\lambda}\right)^k}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

The likelihood function ($L_{\hat{x}}(\lambda, k)$):

$$\begin{aligned} L_{\hat{x}}(\lambda, k) &= \prod_{i=1}^{n} \frac{k}{\lambda}\left(\frac{x_i}{\lambda}\right)^{k-1} e^{-\left(\frac{x_i}{\lambda}\right)^k} \\ &= \left(\frac{k}{\lambda}\right)^n \left(\frac{1}{\lambda^{k-1}}\right)^n \left(\prod_{i=1}^{n} x_i^{k-1}\right) \left(e^{-\sum_{i=1}^{n}\left(\frac{x_i}{\lambda}\right)^k}\right) \\ &= \frac{k^n}{\lambda^{nk}} e^{-\sum_{i=1}^{n}\left(\frac{x_i}{\lambda}\right)^k} \prod_{i=1}^{n} x_i^{k-1} \end{aligned}$$

The log-likelihood function ($\ln L_{\hat{x}}(\lambda, k)$):

$$\ln L_{\hat{x}}(\lambda, k) = n\ln(k) - nk\ln(\lambda) - \sum_{i=1}^{n}\left(\frac{x_i}{\lambda}\right)^k + (k-1)\sum_{i=1}^{n}\ln x_i$$

Partial derivative of the log-likelihood function with respect to λ:

$$\frac{\partial \ln L_{\hat{x}}(\lambda, k)}{\partial \lambda} = -\frac{nk}{\lambda} + k\sum_{i=1}^{n}\frac{x_i^k}{\lambda^{k+1}}$$

Solving for the desired parameter of λ by setting the partial derivative equal to zero:

$$\begin{aligned}
\frac{\partial \ln L_{\hat{x}}(\lambda, k)}{\partial \lambda} &= 0 \\
-\frac{nk}{\lambda} + k \sum_{i=1}^{n} \frac{x_i^k}{\lambda^{k+1}} &= 0 \\
-\frac{nk}{\lambda} + \frac{k}{\lambda} \sum_{i=1}^{n} \left(\frac{x_i}{\lambda}\right)^k &= 0 \\
-n + \sum_{i=1}^{n} \frac{x_i^k}{\lambda^k} &= 0 \\
\frac{1}{\lambda^k} \sum_{i=1}^{n} x_i^k &= n \\
\frac{1}{n} \sum_{i=1}^{n} x_i^k &= \lambda^k
\end{aligned}$$

Therefore the estimator $\hat{\lambda}$ is:

$$\hat{\lambda} = \left(\frac{1}{n} \sum_{i=1}^{n} x_i^k\right)^{\frac{1}{k}}$$

Plugging in $\hat{\lambda}$ into the log-likelihood function ($\ln L_{\hat{x}}(\lambda, k)$) and then differentiating with respect to k in order to find the estimator $\hat{k}$:

$$\begin{aligned}
\frac{\partial \ln L_{\hat{x}}(\lambda, k)}{\partial k} &= \frac{\partial}{\partial k}\left[n \ln k - nk \ln \lambda - \sum_{i=1}^{n} \left(\frac{x_i}{\lambda}\right)^k + (k-1) \sum_{i=1}^{n} \ln x_i\right] \\
&= \frac{\partial}{\partial k}\left[n \ln k - nk \ln \left[\left(\frac{1}{n} \sum_{i=1}^{n} x_i^k\right)^{\frac{1}{k}}\right] - \frac{\sum_{i=1}^{n} x_i^k}{\left[\left(\frac{1}{n} \sum_{i=1}^{n} x_i^k\right)^{\frac{1}{k}}\right]^k} + (k-1) \sum_{i=1}^{n} \ln x_i\right] \\
&= \frac{\partial}{\partial k}\left[n \ln k - \frac{nk}{k} \ln \left(\frac{1}{n} \sum_{i=1}^{n} x_i^k\right) - \frac{\sum_{i=1}^{n} x_i^k}{\left(\frac{1}{n} \sum_{i=1}^{n} x_i^k\right)} + (k-1) \sum_{i=1}^{n} \ln x_i\right] \\
&= \frac{\partial}{\partial k}\left[n \ln k - n \ln \left(\frac{1}{n} \sum_{i=1}^{n} x_i^k\right) - n + (k-1) \sum_{i=1}^{n} \ln x_i\right] \\
&= \frac{n}{k} - \left(\frac{n \sum_{i=1}^{n} x_i^k \ln x_i}{\sum_{i=1}^{n} x_i^k}\right) + \sum_{i=1}^{n} \ln x_i \\
&= \frac{1}{k} - \left(\frac{\sum_{i=1}^{n} x_i^k \ln x_i}{\sum_{i=1}^{n} x_i^k}\right) + \frac{1}{n} \sum_{i=1}^{n} \ln x_i
\end{aligned}$$

Solving for the desired parameter of k by setting the partial derivative equal to zero:

$$
\begin{aligned}
\frac{\partial \ln L_{\hat{x}}(\lambda, k)}{\partial k} &= 0 \\
\frac{1}{k} - \left(\frac{\sum\limits_{i=1}^{n} x_i^k \ln x_i}{\sum\limits_{i=1}^{n} x_i^k} \right) + \frac{1}{n} \sum_{i=1}^{n} \ln x_i &= 0 \\
\left(\frac{\sum\limits_{i=1}^{n} x_i^k \ln x_i}{\sum\limits_{i=1}^{n} x_i^k} \right) - \frac{1}{n} \sum_{i=1}^{n} \ln x_i &= \frac{1}{k}
\end{aligned}
$$

Therefore the estimator $\hat{k}$ is:

$$
\hat{k} = \left[\frac{\sum\limits_{i=1}^{n} x_i^k \ln x_i}{\sum\limits_{i=1}^{n} x_i^k} - \frac{1}{n} \sum_{i=1}^{n} \ln x_i \right]^{-1}
$$